Using Collar-worn Sensors to Forecast Thermal Strain in Military Working Dogs

James R. Williamson, Austin R. Hess, Christopher J. Smalt, Delsey M. Sherrill, and Thomas F. Quatieri
MIT Lincoln Laboratory
Lexington, MA 02421

Catherine O'Brien
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760

***Abstract*—Military working dogs (MWDs) are at high risk of heat strain both during training and missions. Body heat in MWD increases due to work, and the primary means for reducing this heat are resting and panting. Body-worn sensors can enable monitoring of work level and respiratory rate in real time. They can thereby provide real-time objective indicators of thermal strain in MWDs. In this paper a system is proposed for using collar-worn accelerometer, global positioning system (GPS), and audio recorder sensors to provide real-time estimates of work level and respiration (breathing and panting) rate. Automated methods are demonstrated for using a collar-worn accelerometer and GPS sensor to estimate work levels during multiple short-duration activities, and for estimating respiration rates from a collar-worn audio recorder. The potential utility of these estimates for forecasting and monitoring thermal strain is assessed based on performance in out of sample prediction of core temperature (Tc) statistics, which are obtained from ingestible sensors. Using cross-validation, regression models are trained from accelerometer- and GPS-based activity estimates to predict rate of change in Tc, obtaining a correlation of *r=0.59* between actual and predicted Tc change rates. Regression models are also trained from audio-based respiration rate estimates during recovery to predict the Tc values immediately prior to recovery, obtaining a correlation of *r=0.49* between actual and predicted Tc.**

I. Introduction

Military Working Dogs (MWDs), working in a team with a handler, are susceptible to heat illness and injury when working in hot conditions in garrison, training, and theater environments [1-5]. Heat stroke was the third most common cause of death in MWDs during conflicts in Iraq and Afghanistan, and was the most common cause of death from non-battle injury [4]. Outside of combat, heat injury is the second most common reason (behind behavioral causes) for discharge from duty of younger (<5 years) MWDs [5].

To develop tools for mitigating risk associated with excessive thermal strain, an understanding of baseline data regarding the excursions in MWD body temperature during normal activities is needed. Providing MWD handlers the ability to monitor MWD thermal status in real-time could minimize the risk of heat injury during exposure to thermal extremes while working, training, and resting.

Dangerously high core temperature (Tc) typically arise in MWDs during work. This work can be estimated using body movement measures obtained from collar-worn accelerometer and GPS sensors. Heat loss and recovery from high Tc values are typically accomplished in MWDs by resting and panting [6-8]. Resting is easily estimated from the movement sensors, and respiratory rates including panting (4-6 Hz) are estimated using a collar-worn audio recorder.

In this paper we demonstrate methods for automated estimation of activity levels from collar worn movement sensors and of respiration rates from a collar worn audio recorder. The methods are applied to a field data set collected over two days from six MWDs conducting several military-relevant activities. The Tc of MWDs was measured using ingestible telemetric temperature sensors. Movement and respiration estimates are assessed using statistical models to predict Tc and ΔTc during activity.

The field data collection is described in Section II. Feature extraction and statistical modeling algorithms are presented in Section III. Results are presented in Section IV, showing the ability to predict Tc and ΔTc values on out of sample test data. Finally, a summary of the work and future directions are given in Section V.

Distribution A: Public Release.This work was sponsored by the United States Army Research Institute of Environmental Medicine under United States Air Force Contract FA8721-05-C-0002. The opinions or assertions contained herein are the private views of the author(s) and are not to be construed as official or reflecting the views of the Army or the Department of Defense.

In conducting the research described in this report, the investigators adhered to the "Guide for the Care and Use of Laboratory Animals" as prepared by the Committee for the Update of the Guide for the Care and Use of Laboratory Animals of the Institute for Laboratory Animal Research, National Research Council.

Any citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement of approval of the products or services of these organizations.

II. Data Collection

The 341st Training Squadron at Lackland Air Force Base (AFB), San Antonio, TX, procures and trains MWDs for patrol and detection activities, spending approximately 60 days in each phase of the training [9]. Once certified, MWD teams support military operations and their capabilities are used to enhance the effectiveness of the unit. The MWD training includes the following skills.

- **Obedience:** Sit, Down, Heel, Stay.
- **Obstacle Course:** Negotiate obstacles under control.
- **Controlled Aggression:** Pursue, bite and hold, Out, Heel, Stay.
- **Building Search:** Detect a person, narcotics or explosives hidden inside a building.
- **Vehicle Search:** Detect narcotics or explosives hidden on a vehicle.
- **Scouting:** Locate a person by scent, sight or sound.
- **Gunfire:** Remain calm and attentive while weapons are fired.

A total of forty-eight MWDs participated in this study, which was approved by the Department of Defense MWD Veterinary Service (DODMWDVS) Institutional Animal Care and Use Committee (IACUC). Twenty-four MWDs in training for the above skills were tested, six MWDs at a time, over two consecutive days on each of four occasions (August and October 2014, March and August 2015). Another twenty-four MWDs that were not in training at the time were tested in July 2015 during a single exercise walk around the kennel grounds.

This paper is a preliminary analysis of the data measured from six MWDs that were tested on October 22 and 23, 2014. These MWDs were tested on three skills per day, with each test separated by long rest periods. The Oct. 22 tests were *Obedience*, *Building Search*, and *Aggression*, and the Oct. 23 tests were *Obstacle Course*, *Scouting*, and *Gunfire*. The test durations ranged from 3.5 min. to 19.8 min., with an average duration of 9.4 min. The average meteorological conditions during the testing (WeatherHawk 500, Logan, UT) are shown in Table 1.

Table 1. Average meteorological conditions.

Time	0800	1100	1300
Temperature ^{o}C	20	25	27
Relative Humidity %	65	49	73
Solar Radiation W/m^2	45	400	567
Wind Speed m/s	0.4	1.8	1.3

MWD instrumentation consisted of the following. An ingested telemetric temperature pill (VitalSense, Phillips Respironics, Murrrysville, PA) was used to obtain core body temperatures, transmitted to a body-worn data logger (Equivital™, Hidalgo Ltd., Cambridge, UK). This data logger provided Tc measurements only during the test intervals, while it was being worn. In this paper, only these Tc measurements are used. Additional Tc measurements, obtained from an off-body data logger (VitalSense®, Mini Mitter, Bend OR) during pre-test and post-test intervals, will be used in future data analyses.

A three-axis accelerometer (Actigraph, Pensacola, FL), and a GPS device (Garmin Foretrex, Garmin, Ltd., Olathe, KS) were worn on the collar during the tests. Finally, a Zoom H1 (Zoom North America, Ronkonkoma, NY) audio recorder was located on the collar close to the dog's mouth in order to detect breathing and panting. Additionally, off-body voice recorder data is also available both prior to and following each test, allowing the monitoring of respiration during those periods. Fig. 1 illustrates how the accelerometer and audio recorder devices were attached to a dog collar, as well showing typical data obtained from these devices.

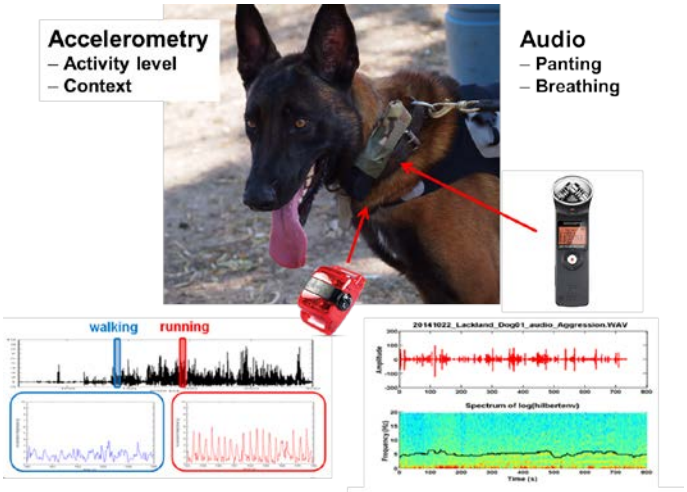

Fig.1. Collar-worn accelerometer and audio recorder sensors used to estimate activity level and respiratory rate.

III. Methods

A. Feature Extraction

There are two primary goals of our study: 1) we seek to demonstrate that a collar-worn accelerometer and GPS sensor can be used to estimate the activity levels in MWDs, and that these estimates have real world utility in terms of predicting Tc increases; 2) we seek to demonstrate that a collar-worn audio recorder can be used to estimate respiratory rates during recovery, and that these estimates have real world utility in terms of reflecting the level of heat strain (i.e., Tc) achieved by the MWD during work.

1) *Activity level from accelerometer and GPS*

To estimate work from an accelerometer and GPS, measures are obtained from short (5 s) data frames. These measures are integrated over time to obtain cumulative work estimates, which can be used to predict cumulative increase in core temperature over an entire test.

Accelerometer-based activity measures are obtained at each time step (100 Hz) from the magnitude of the three-dimensional acceleration vector, a_i, which is independent of sensor orientation. Next, the variance of a_i is computed over the j^{th} frame,

$$b_j = \mathrm{var}\left(a_{i \in f_j}\right). \quad (1)$$

Finally, the 5 s frame-level b_j measures are summed across all frames in the k^{th} test, thereby obtaining an integrated accelerometry-based activity measure,

$$c_k = \sum_{j \in g_k} b_j . \quad (2)$$

Fig. 2 plots cumulative c_k values obtained during each of the six skill tests from Dog 1. Observe that there are considerable difference in activity durations and cumulative activity profiles.

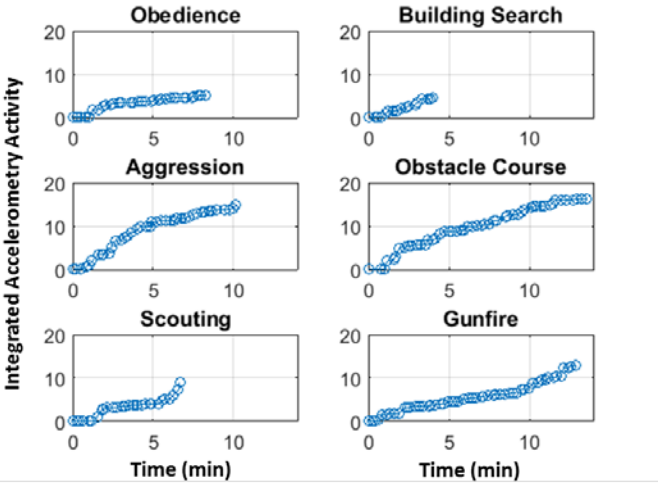

Fig.2. Integrated activity levels estimated from accelerometer during the six skill tests of Dog 1.

GPS-based activity measures are obtained using a similar procedure. The GPS speeds, x_i, are computed at one second intervals, and then the variance of the GPS speed measures are computed within the j^{th} frame,

$$y_j = \mathrm{var}\left(x_{i \in f_j}\right). \quad (3)$$

Finally, the 5 s frame-level y_j measures are summed from all frames within the time interval of the k^{th} test, thereby obtaining an integrated GPS-based activity measure,

$$z_k = \sum_{j \in g_k} y_j . \quad (4)$$

Fig. 3 plots cumulative z_k values obtained during each of six tests from Dog 1. Observe that there are rough similarities between these GPS-based activity estimates and the accelerometer-based estimates in Fig. 2. Fig. 4 plots measured Tc values during the same six tests for which activity level estimates are plotted in Figs 2 and 3.

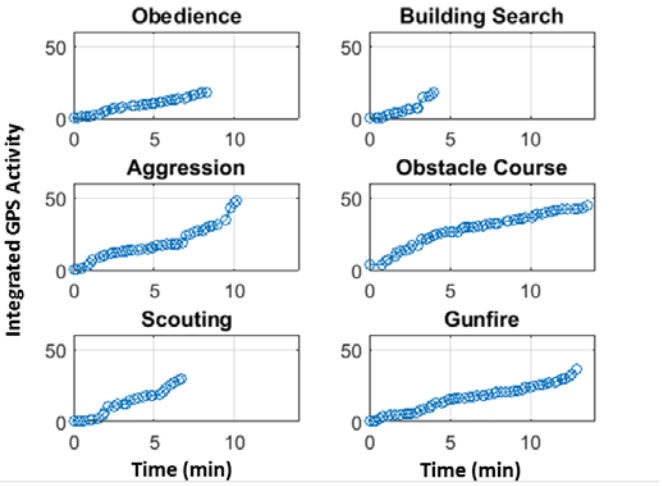

Fig.3. Integrated activity levels estimated from GPS during six tests of Dog 1.

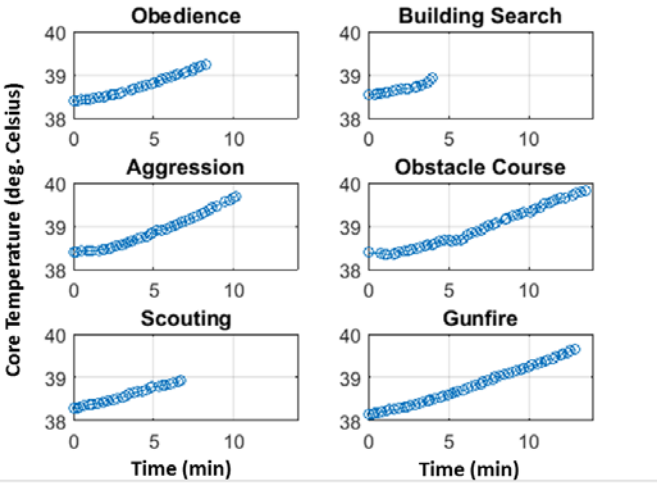

Fig.4. Core temperature (Tc) profiles measured during six tests of Dog 1.

2) Respiratory rate (RR) from audio

Respiratory rate (RR) estimates are computed from the audio signal based on periodicities in the respiratory frequency range. The estimation algorithm, developed independently of the regression experiments described in Section IIIB, obtains RR estimates in a three stage process: 1) extracting the energy envelope from the audio signal, 2) generating breathing period hypotheses from the energy envelope for each 3 s frame (with 1 s overlap between successive frames), and 3) selecting valid RRs based on the likelihoods of RR hypotheses.

Extracting energy envelope: A pre-emphasis filter that amplifies lower frequency components is applied, followed by band-pass filtering from 500-1400 Hz. Next, the signal is downsampled by a factor of 10, and finally the log-magnitude of the Hilbert envelope of the waveform is computed [10].

Generating respiratory period hypotheses: From the log Hilbert energy envelope, multiple hypotheses are generated within local time frames. Three-second time frames are used,

with one-second overlap between successive frames. Within each frame, candidate respiratory period peaks are obtained from an autocorrelation function of the energy envelope, constrained to an allowable range of 0.1-1.0 seconds (spanning possible respiratory rates of 1-10 Hz). Respiratory period hypotheses are obtained from local peaks in the autocorrelation function. In order to make these peak hypotheses more robust, the following steps are first performed.

- **Biasing in favor of short respiratory periods:** the autocorrelation function of the log Hilbert envelope is modulated by a linear function that ramps from a value of one at zero seconds delay down to a value of zero at one seconds delay. This is done to favor selection of the shortest respiratory period supported by the data.
- **Smoothing across time:** smoothing within each autocorrelation delay bin across nearby frames (Gaussian kernel, $\sigma = 4$ s).
- **Smoothing across delays:** smoothing over autocorrelation delay bins within each frame (Gaussian kernel, $\sigma = 0.01$ s).

Selecting a single valid respiratory rate per frame: For each frame (with one frame per second), a single breathing period is selected based on the local context. This is done by using a Gaussian mixture model (GMM) [11] to assign likelihoods to each respiratory period hypothesis, based on the context of other hypotheses nearby in time. Specifically, it uses a one-dimensional feature vector containing the respiratory period values over a local ±30 s neighborhood. For each target frame, a three-component GMM forms clusters around the breathing period hypotheses that garner the most contextual support. The likelihood assigned to each respiratory period hypothesis is based on its GMM likelihood multiplied by the magnitude of its log Hilbert envelope autocorrelation peak.

This procedure is repeated for three iterations. In the first iteration, there are three candidate respiratory periods per frame. In iterations 2 and 3, there is only one candidate per frame. This causes the GMM to converge onto a single hypothesis that is consistent with evidential support in the local temporal neighborhood. The final likelihood score obtained for the winning hypothesis in each frame is used to determine the validity of the hypothesis: it is valid only if the likelihood exceeds a threshold of 0.0001. The respiratory rate (RR), in units of Hz, is computed as the reciprocal of the selected respiratory period in each frame.

Fig. 5 plots RR estimates corresponding to the Controlled Aggression test for Dog 1. The black lines delineate the test interval, with the pre-test interval to the left and the post-test interval to the right. The valid RR estimates are plotted in red and the invalid estimates in blue. Observe that the RR rate quickly climbs during the post-test interval into the 4-6 Hz panting range, with panting persisting throughout the measured recovery interval. This behavior may be attributable to the relatively high Tc (39.7 deg.) measured at the end of the Aggression test (see the rightmost points in Fig. 4).

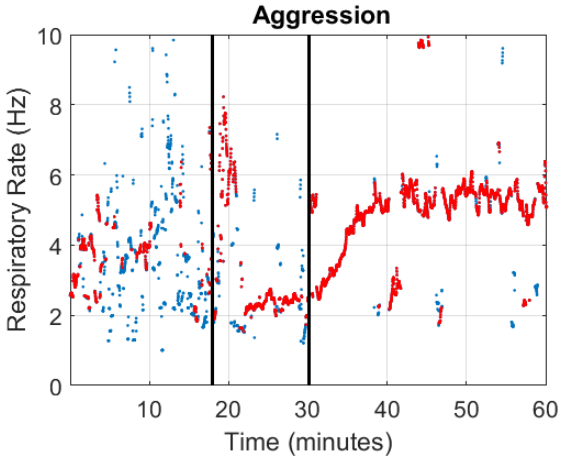

Fig.5. Respiratory rate (RR) estimates obtained from pre-test, test, and post-test intervals. Estimates labels as valid estimates are plotted in red.

Fig. 6 illustrates the RR estimates obtained for the Obedience test, which culminates in a Tc (39.2 deg.) that is half a degree lower than the Tc at the end of the Aggression test (see Fig. 4). Observe that the Obedience recovery period contains RR values that alternate between panting and regular breathing. Together, Figs 5 and 6 suggest that a higher Tc at the end of a test results in more panting during recovery.

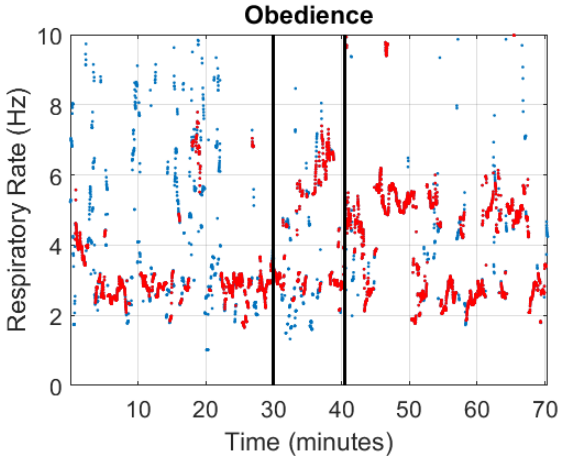

Fig.6. Respiratory rate (RR) estimates obtained from pre-test, test, and post-test intervals. Valid estimates are plotted in red.

B. Statistical Modeling and Prediction

In order to assess the real world utility of our features, statistical models are constructed that 1) map the activity level features to Tc change rates (ΔTc/min), and 2) map the recovery-interval RR features to the Tc values immediately prior to recovery. Using cross-validation, the utility of the activity and RR-based features for predicting Tc statistics is assessed using Gaussian staircase regression (G-SR) statistical models. G-SR has previously been applied to predicting severity of major depressive disorder from speech and video [12-13], severity of Parkinson's disorder from speech [14], and carried load levels from accelerometry [15].

G-SR accomplishes multivariate regression via two steps: 1) multivariate fusion, and 2) univariate regression. Multivariate fusion is done by constructing several partitions of the outcome variable, with each partition being used to train a different Gaussian classifier. The set of Gaussians belonging to the lower partitions are combined into a single ensemble of Gaussians representing Class 1, and the set of Gaussians belonging to the higher partitions are combined into an ensemble representing Class 2. The log-likelihood ratios (LLRs) of the two classes obtained from these Gaussian ensembles on the training set are then used to construct a least-squares univariate regression model mapping the LLRs into the outcome variable. This model is used to generate a prediction of the outcome variable based on the LLR from test data.

1) *Predicting ΔTc from activity features*

There are 36 test trials (six different MWDs with six tests per MWD). Since the tests vary in duration, it is appropriate to predict the average rate of change in Tc in each test, based on the extracted activity features. For each test trial, a statistical model mapping the average activity level into the average rate of change in Tc, ΔTc/min, is constructed from the training data. Over all 36 trials, the outcome variable (ΔTc/min) ranges in value from 0.03 to 0.24. Using G-SR, outcome variable delimiters of {0.06, 0.10, 0.14} are used to construct an ensemble of four Gaussian distributions each for Class 1 (lower ΔTc/min) and Class 2 (higher ΔTc/min). The features, acceleration-based activity and GPS-based activity, are first z-scored based on the training set. Feature regularization is done by adding σ^2=100 to the diagonal elements of the G-SR covariance matrices. During testing, the G-SR Gaussian ensembles produce a LLR for each trial, which is mapped to the outcome variable (ΔTc/min) via a univariate linear regression model, which is constructed using least-squares fitting based on the training set LLRs.

2) *Predicting Tc from RR features*

Audio data is available from only 28 of the 36 trials. Our goal is to ascertain how well RR features obtained during the post-activity recovery interval (e.g., rightmost interval of Figs. 5 and 6) can predict the Tc values at the end of the skill tests (e.g., rightmost points in Fig. 4). Note that, in our current analysis, Tc measurements obtained during the skill tests only are used. In future work additional Tc measurements obtained from the VSP logger during recovery will be analyzed as well. For each trial, a statistical model is constructed from the other 27 trials mapping the RR features into the final Tc value measured prior to the recovery interval. From all 28 trials, the outcome variable Tc ranges in value from 38.7 to 40.8 deg. Celsius. G-SR partitions the training data set based on outcome variable delimiters of [39.0:0.1:40.0], thereby constructing an ensemble of 11 Gaussian distributions each for Class 1 (lower Tc values) and Class 2 (higher Tc values).

RR estimates over the recovery interval vary along two completely different dimensions: 1) the prevalence of RR values designated as being *valid*, and 2) the average value of the valid RR estimates. For example, in comparing the pre-activity and post-activity intervals in Fig. 5, we can see time regions where the average values of valid RR estimates are similar, but where the prevalences of RR validity are very different. Therefore, it is necessary to capture both feature dimensions in order to disassociate high-frequency RRs obtained with low prevalence from those obtained with high prevalence. To capture both dimensions we construct RR features based on the fraction of valid RRs that are above a given threshold value. In order to also capture the distributional properties of valid RRs, we sweep over a range of thresholds spanning the breathing and panting frequencies, [1.125:0.125:5.875], and thereby produce a 39-dimensional vector of joint prevalence-frequency features. Fig. 7 plots the RR-based feature vectors obtained using this method from the recovery periods shown in Figs 5 and 6.

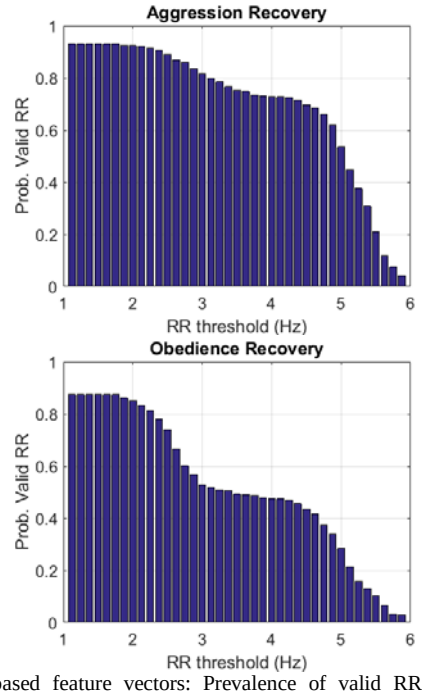

Fig.7. RR-based feature vectors: Prevalence of valid RR estimates as a function of RR thresholds. Top: result for Aggression recovery interval (see Fig. 5), in which Tc = 39.7 deg. Bottom: result for Obedience recovery interval (see Fig. 6), in which Tc = 39.2.

The RR-based features are z-scored based on the training set before being used by G-SR, and feature regularization is done by adding σ^2=1 to the diagonal elements of the G-SR covariance matrices. Least squares linear univariate regression is used to map the LLRs to the outcome variable (Tc).

IV. RESULTS

1) *Predicting ΔTc/min from activity features*

The accelerometry- and GPS-based activity features are used to predict ΔTc/min values from all 36 trials. This is done by

applying G-SR within cross-validation, resulting in correlations between predicted and true ΔTc values of *r=0.54* ($p<0.001$) for accelerometry-based activity estimates, *r=0.35* ($p<0.05$) for GPS-based activity estimates, and *r=0.59* ($p<0.001$) for combined activity estimates (i.e., combined using two-dimensional feature vectors). Fig. 8 shows a scatter plot of true versus predicted ΔTc/min for the combined result, along with a linear regression fit.

2) Predicting Tc from respiratory features

The 39-dimensional RR-based feature vectors are used to predict Tc values prior to recovery from the 28 trials for which audio data is available. This is done by applying G-SR within cross-validation, resulting in correlation of *r=0.49* ($p<0.01$) between predicted and actual Tc values (Fig. 9).

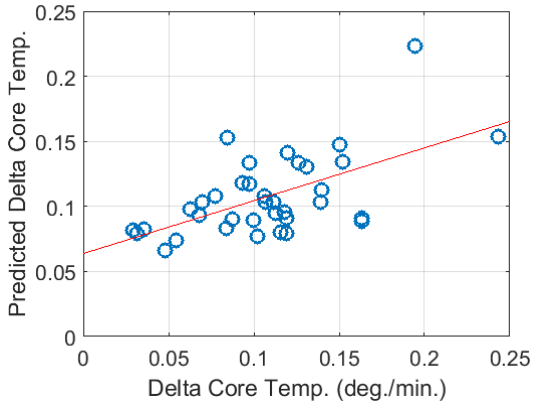

Fig.8. Predicted ΔTc/min from all 36 trials as a function of true ΔTc/min, obtained using both accelerometry- and GPS-based activity estimates.

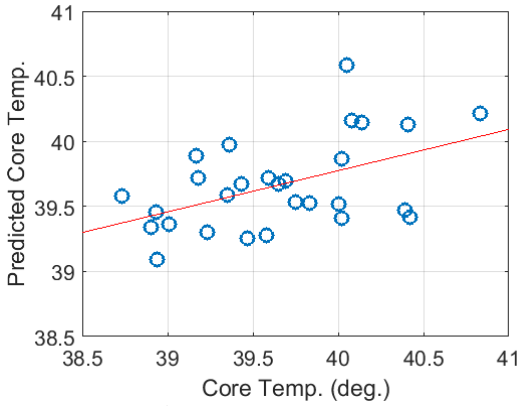

Fig.9. Predicted Tc values from the 28 trials that contain audio as a function of true Tc. Predictions were obtained using the 39-dimensional RR-based feature vectors (see Fig. 7),

V. CONCLUSION

There is a vital need for real-time estimates of risks for thermal strain in MWDs, which could influence handlers' decisions regarding work and rest cycles and improve health and safety outcomes for these valuable animals. This paper shows that it is possible to obtain useful estimates of activity levels and respiratory rates from collar-worn sensors. Using out-of-sample testing we showed that these estimates can inform predictions of change in Tc (from activity) and Tc levels (from RR). Moderately strong correlations between predicted and actual Tc values were obtained despite limitations due to the small size of the training set, the short durations of the MWD skill tests, and the large variety of behaviors comprising the skill tests. We plan to expand our analysis based on the entire set of MWD data already collected, as well as future planned collections. The new analysis, which will critically include Tc measurements during recovery, will hopefully lead to improvements in our ability for monitoring and forecasting the risk of heat strain in MWDs.

REFERENCES

[1] Robinson, F. R., & Garner, F. M. (1972). Histopathologic survey of 2,500 German Shepherd military working dogs. ARMED FORCES INST OF PATHOLOGY WASHINGTON DC.

[2] Bruchim, Y., Klement, E., Saragusty, J., Finkeilstein, E., Kass, P., & Aroch, I. (2006). Heat stroke in dogs: a retrospective study of 54 cases (1999–2004) and analysis of risk factors for death. *Journal of veterinary internal medicine*, *20*(1), 38-46.

[3] Toffoli, C. A., & Rolfe, D. S. (2006). Challenges to military working dog management and care in the Kuwait theater of operation. *Military medicine*, *171*(10), 1002-1005.

[4] Miller, L. Causes of combat mortality of military working dogs in OIF and OEEF, 2007-2012: A piolot study of 109 cases. Raleigh, NC: Canine Science and Technology Workshop; 2014.

[5] Evans, R. I., Herbold, J. R., Bradshaw, B. S., & Moore, G. E. (2007). Causes for discharge of military working dogs from service: 268 cases (2000–2004). *Journal of the American Veterinary Medical Association*, *231*(8), 1215-1220.

[6] Crawford, E. C. (1962). Mechanical aspects of panting in dogs. *Journal of applied physiology*, *17*(2), 249-251.

[7] Schmidt-Nielsen, K., Bretz, W. L., & Taylor, C. R. (1970). Panting in dogs: unidirectional air flow over evaporative surfaces. *Science*, *169*(3950), 1102-1104.

[8] Goldberg, M. B., Langman, V. A., & Taylor, C. R. (1981). Panting in dogs: paths of air flow in response to heat and exercise. *Respiration physiology*, *43*(3), 327-338.

[9] Military Working Dog Program. Washington, DC: Department of the Army; 2013. Report No.: AR 190-12.

[10] Molla, M. K. I., & Hirose, K. (2007). Single-mixture audio source separation by subspace decomposition of Hilbert spectrum. *Audio, Speech, and Language Processing, IEEE Transactions on*, *15*(3), 893-900.

[11] Bishop, C. M. (2006). Pattern Recognition. *Machine Learning*.

[12] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Horwitz, R., Yu, B., & Mehta, D. D. (2013, October). Vocal biomarkers of depression based on motor incoordination. In *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge* (pp. 41-48). ACM.

[13] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., & Mehta, D. D. (2014, November). Vocal and Facial Biomarkers of Depression Based on Motor Incoordination and Timing. In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge* (pp. 65-72). ACM.

[14] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Perricone, J., Ghosh, S. S., Ciccarelli, G., & Mehta, D. D. (2015). Segment-dependent dynamics in predicting Parkinson's disease. In *Sixteenth Annual Conference of the International Speech Communication Association*.

[15] Williamson, J. R., Dumas, A., Ciccarelli, G., Hess, A. R., Telfer, B. A., & Buller, M. J. (2015, June). Estimating load carriage from a body-worn accelerometer. In *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on* (pp. 1-6). IEEE.